Toward Humanoid Choreography and Dance

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Da Che B.S.

Graduate Program in Electrical and Computer Science

The Ohio State University

2012

Master's Examination Committee:

Hooshang Hemami, Advisor

Yuan F. Zheng
Abstract

A humanoid robot that can perform choreographic movements and dance to music is considered. The robot predicts the beat and intensity of the music, then it performs learned movements. A music analysis algorithm is considered to extract and predict beats and music intensity from audio signals. The extracted music features are used to synthesize the dance pattern. A two-link robot that would create movement and perform dance when it hears the music is envisioned. A position and velocity feedback control system guides the robot in the dance pattern. Computer simulations are performed to demonstrate the feasibility of the method.
Dedication

This document is dedicated to my family.
Acknowledgments

I would like to express my gratitude to my advisor, Professor Hooshang Hemami, whose expertise, understanding, and patience, added considerably to my graduate experience.
Vita

August 2003 ..................................................Dalian No. 24 High School, China

September 2006 ...........................................B.S. Electrical Engineering and Automation,

Tianjin University, China

September 2010 ..........................................M.S. Department of Electrical and

Computer Engineering, The Ohio State

University

Fields of Study

Major Field: Electrical and Computer Engineering
Table of Contents

Toward Humanoid Choreography and Dance .......................................................... 1

THESIS ...................................................................................................................... 1

Abstract .................................................................................................................. ii

Dedication .............................................................................................................. iii

Acknowledgments ................................................................................................. iv

Vita .......................................................................................................................... v

Fields of Study ....................................................................................................... v

Table of Contents ................................................................................................. vi

List of Tables ......................................................................................................... vii

List of Figures ......................................................................................................... ix

Chapter 1: Introduction .......................................................................................... 1

Chapter 2: MUSIC FEATURE ANALYSIS .......................................................... 3

2.1 General Description of the Beat Prediction Algorithm .................................... 3

2.2 Pre-processing of the Audio Signals ................................................................. 7

2.3 Tempo and Phase Prediction .......................................................................... 9

2.4 Steady State Error ........................................................................................ 14
<table>
<thead>
<tr>
<th>Section/Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Music Intensity Analysis</td>
<td>15</td>
</tr>
<tr>
<td>Chapter 3: Simulation</td>
<td>16</td>
</tr>
<tr>
<td>3.1 Result of the Beat Predicting Algorithm</td>
<td>16</td>
</tr>
<tr>
<td>3.2 Robot Choreography</td>
<td>22</td>
</tr>
<tr>
<td>Chapter 4: Conclusions and Future Work</td>
<td>26</td>
</tr>
<tr>
<td>References</td>
<td>27</td>
</tr>
<tr>
<td>Appendix A: Model of the Two-Link Robot</td>
<td>31</td>
</tr>
<tr>
<td>Appendix B: Relations between Two Beat Prediction Algorithms</td>
<td>38</td>
</tr>
</tbody>
</table>
List of Tables

Table 1 Parameters used in the Two-Link robot model......................................................... 36
List of Figures

Figure 1 two of the locally generated signals which consist of rectangular pulses at different frequencies and with different phases. .......................................................... 5
Figure 2 Flowchart of Main Steps the Beat Prediction Algorithm ....................... 6
Figure 3 Details about onset extraction. ........................................................................ 8
Figure 4 Onset Extraction .......................................................................................... 17
Figure 5 Tempo estimates using the algorithm. ......................................................... 19
Figure 6 Phase estimates using the algorithm .......................................................... 21
Figure 7 the Two-Link robot model moves to the reference input θ1 and θ2 .......... 23
Figure 8 Flowchart of the system ............................................................................. 25
Figure 9 Two-Link Model of the Robot .................................................................... 31
Figure 10 Flowchart of the control part ................................................................. 35
Figure 11 Response curve of θ to step input ........................................................... 37
Humanoid robots that behave like humans have been subject of recent studies [12], [27]. These humanoid Robots can walk, run, and perform more complex whole body movements [12], [16]. If robots have intelligence and shapes similar to those of humans, it is possible that humanoid robots can communicate with humans by other means, such as gestures, dance, play and sports [27]. Dance, performed by humanoid robots, can be enjoyed as a kind of entertainment and art [16]. Humanoids can share a common brain and, therefore, collaborate in structurally complex dance that currently is programmed by humans primarily in pairs.

Many researchers are involved in improving the physical performance of robots [8] [12], [20]. Sony robot QRIO can produce many dance movements [26]. It has been involved in a learning environment [6] for studying the interaction between a human and a robot. This robot is also capable of reproducing some partner's movements. Kosuge et. al. [30] exhibited a robot dancer (called MS DanceR) which danced with a human partner. Most of the motion of these robots is pre-programmed, and not in response to audio signals.

Kazuyoshi et. al. [17] developed a biped robot that could keep step to music. They also invented a robot that could sing to the music [15]. However, it takes about 25 seconds to synchronize their movement to the music [17]. Another humanoid robot HRP-4C [14] has realistic human head and body based on a young Japanese female. This robot has
presented a very impressive dance performance at Tokyo's Digital Content EXPO 2010 [25].

As described by Shiratori et al. [28], the ability to dance to music is a natural born skill for a human. They also point out the strong connection of dance with music in two ways: both the rhythm and the intensity of the dance motion are synchronized with the music. The first assumption is based on the fact that almost all people can recognize the rhythm of music. The second assumption is based on the fact that people feel relaxed while listening to soft music and feel excited while listening to agitated music.

Based on the assumptions mentioned above, a beat prediction algorithm is considered for choreography and dance in the paper. The goal is to develop a humanoid robot that can perform and dance like human dancers. A two-dimensional two-link robot model could implement some simple dancing movements. In addition, the robot must synchronize to the music. As the first step toward this goal, a two-dimensional two-link model of a human arm is considered in this paper.

The remainder of this paper is organized as follows: Section 2 introduces the music feature analysis. Section 3 presents the simulation of the robot moving to music. Section 4 summarizes the study.
Chapter 2: MUSIC FEATURE ANALYSIS

This section introduces the intelligence of the robot dancer. Two features of the music: rhythm and musical intensity are analyzed here. To detect the rhythm of the music, a real-time beat prediction algorithm is presented. The outputs of the algorithm are the tempo and phase of the beats. The main idea of the algorithm is introduced first. The necessary pre-processing is presented next. Details about how to detect and predict the tempo and period of the music are discussed next. Subsequently, the results of the algorithm are modified for more precision. Finally, a musical intensity method is introduced.

2.1 General Description of the Beat Prediction Algorithm

The beat is the basic unit of time in music, the pulse of the mensural or beat level [1], [29]. The extracted beat signal is a periodic pulse signal with two components: a frequency and a phase relative to a given time reference. The frequency or the tempo of the beat means the number of pulses per minute. The phase of the beat indicates where the beat occurs during the period [4]. In the beat level, pulses are heard as the basic time unit of the piece [21]. In many cases, the beats one hears are the quarter notes. In most of the music, beats usually are accompanied with instruments like drums, the guitar, and the piano and have larger energy [4]. People who have a good sense of music can extract and predict the beats because they can distinguish the instruments played at quarter notes.
In our beat prediction algorithm, beats are detected and predicted by counting the quarter notes. The counting method is to some degree similar to what humans do. Once the quarter notes with large energy are extracted, the algorithm can calculate the tempo and phase of the beats. It also can predict the beats in the near future.

The algorithm is briefly described here, and details are given in the following sections.

First of all, a class of signals is locally generated in time domain as shown in figure 2-1. Signals are periodic but with different frequencies, consisted of rectangular pulses, and lasts 10 seconds.

This class of signals is generated in two steps:

Step one: a group of periodic, ten seconds long signals with different periods are generated. Periods of the signals are equally distributed form 0.2 second to 2.0 second.

Although the length of period in each signals are different, each period in every signal has a rectangular pulses at the beginning of the period, as shown in Figure 1(a).

Step two: with respect to each one of signals generated in step one, a group of signals are generated with different delays. Delays are equally distributed from zero to the length of the period. Signals are truncated once they are longer than 10 seconds, as shown in Figure 1(b).
Figure 1: two of the locally generated signals which consist of rectangular pulses at different frequencies and with different phases.

So far, a class of locally generated signals with different periods and delays which consists of rectangular pulses is generated. The envelopes of these signals are similar to music beats with different periods and phases. For the detection of the frequency and the phase, the audio signal is compared with every one of the locally generated signals. The period and delay of the locally generated signal which has the greatest similarity to the audio signal are selected to be the period and phase of the music.
In the algorithm, some ideas and refinements are introduced to achieve better results and simplify the calculations. The flowchart of main steps the beat prediction algorithm is shown in Figure 2 and it is modified from [15].

![Flowchart of Main Steps the Beat Prediction Algorithm](image_url)

**Figure 2 Flowchart of Main Steps the Beat Prediction Algorithm**
The method is briefly described here. The audio signal is processed through six band pass filters. The amplitude envelope of the signal is calculated for each filters' output. These signals are differentiated and half-wave rectified to get the onset signals. After that each of the six onset signals is convolved with a group of model signals. These model signals are periodic signals which only consist of unit pulses but have different periods and phases. Then the energies of the results of the convolutions from different frequency bands are summed up. Finally, the maximum point of the summed energy signal is picked and the tempo calculated. Once the tempo and phase of the beats are calculated, the algorithm can use this information to predict future beats.

2.2 Pre-processing of the Audio Signals

In order to get more accurate results and simplify the calculation, the original audio signal is pre-processed before comparison with locally generated signals. Pre-processing procedure includes filtering and onset extraction, as shown in this section below.
As mentioned before, the algorithm focuses only on the tempo and the phase of the beats. However, in a piece of music, there is much information. Usually, the tempo of the beats in a piece of music is between 20 and 300 beats per minute (bpm), while the sample frequency of the audio signal is 44100 Hz. So, most of the information contained in the signal is useless in this circumstance. In order to simplify the calculations and get better results, the unrelated information is removed by a pre-processing step.
A piece of music as a whole has the same rhythm regardless of the frequency band [4]. Since the algorithm is not sensitive to the frequency of the signal, filter banks are used to divide audio signal into six bands for more accuracy. Thus, the algorithm is executed in different frequency bands.

The filter bank used here has six bands. The lowest band is a low pass filter with a cutoff frequency at 200 Hz. The next four bands are band pass filters with cutoffs at 200 Hz and 400 Hz, 400 Hz and 800 Hz, 800 Hz and 1600 Hz, 1600 Hz and 3200 Hz. The highest band is a high pass filter with a cutoff frequency at 3200 Hz. In case one band with strong power may overwhelm other bands, the outputs are normalized.

Each of the filtered signals is converted into digital form for further operations, and sample frequency is 44100 Hz here to avoid distortion. Then, the onset is extracted from each of the six signals. Details about onset extraction process can be divided into three parts as shown in Figure 3. In process of onset extraction, the envelope of the input signal is extracted first. Then the signal is differentiated and half-wave linear rectified.

**Figure 3** Details about onset extraction.
In order to eliminate details of the music, the envelope of the audio signals from the six bands are extracted. This step would barely lose any beat information and simultaneously simplify the calculations. Here the normalized filter bank outputs are convolved with a half-Hanning (raised cosine) window. When processed, the most resent inputs are emphasized and rapid changes masked. So, the signal will be smoothed. The next step is to down sample the smoothed outputs to 200 Hz. The outputs of the down sampled signals are the envelope of the audio signals.

A single beat may last for tens of milliseconds, so it may lead to inaccuracy. In order to minimize this part of error, the onset of beats, or the start point of beats, should be picked. So, after calculating the envelope of the signal, the envelope signals are differentiated and half-wave rectified. Therefore, the onset signals from six frequency bands are extracted from the original audio signal and can be used for further calculation.

2.3 Tempo and Phase Prediction

A method that can calculate tempo and phase of the beats from the onset signal is discussed here. The beats have pulse shape in the onset signal, because the amplitude of the signal is comparatively large. It can be supposed that the onset signal consists of two components: a group of pulses formed because of beats in the original audio signal and the rest of the signal regarded as noise. The noise consists of other information from the audio signal.
and the white noise from recording the music. Once these two parts are separated, the
period and phase of the beats can be easily calculated. However, extracting the pulses
directly from the signal is difficult.

An alternative method to overcome the above difficulty is as follows. A class of locally
generated periodic discrete signals which consists of rectangular pulses as introduced
above is generated in time domain. Each of the locally generated signals has 2000
samples, with a sample frequency of 200 Hz. So, every locally generated signal lasts 10
seconds and the time between two adjacent samples is 0.05 sec.

The class of locally generated signals is generated in two steps. Step one: 380 signals as a
group are generated as follows:

\[
g_{il}[n] = \sum_{k=-\infty}^{\infty} \delta[n - k \cdot T_i] \quad n = 1,2,3 \ldots 2000
\]

\[
T_i = 40 + i \quad i = 1,2,3 \ldots 380 \quad \text{Equation 1}
\]

Where \(n\) is the index of the signal, \(\delta[\cdot]\) is the discrete impulse, \(g_{il}[\cdot]\) is the \(i^{th}\) signal in the
group, and \(T_i\) is the period of the corresponding signal \(g_{il}[\cdot]\) in the samples. Since that the
locally generated signals are discrete and that these signals are used to be compared with
the onset audio signal, discrete impulses are used to replace rectangular pulses mentioned
before.

Step two: with respect to each one of signals \(g_{il}[\cdot]\) generated above, a group of signals are
generated as follows:
Where \( g_{ij}[\cdot] \) is the \( j^{\text{th}} \) signal generated from \( g_i[\cdot] \), and \( D_j \) is the corresponding time delay in samples. The number of signals generated from \( g_i[\cdot] \) numerically equals to corresponding period \( T_i \).

Thus, the class of locally generated signals is defined, and they can be regarded as a class of standard signals of the beats extracted from the music with different periods and phases. Then, each of these locally generated signals should be compared to the onset audio signals. The one with the greatest similarity to the onset audio signals is marked and saved. That signal is regarded as the onset beat signal extracted from the audio signal, and the corresponding period and time delay are regarded as the period and phase of the audio signal. In order to compare the similarity, similarity factor \( S_{ij} \) is defined as follows:

\[
g_{ij}[n] = \sum_{k=-\infty}^{\infty} \delta[n - k \cdot T_i - D_j] \quad n = 1,2,3 \ldots 2000
\]

\[
D_j = j \quad j = 0,1,2,3 \ldots T_i - 1
\]

Equation 2

\[
s_{ijk} = \sum_{n=1}^{2000} y_k[n] \cdot g_{ij}[n]
\]

\[
S_{ij} = s_{ij1} + s_{ij2} + s_{ij3} + s_{ij4} + s_{ij5} + s_{ij6}
\]

\[
i = 1,2,3 \ldots 380; \quad j = 1,2,3 \ldots T_i - 1; \quad k = 1,2,3 \ldots 6
\]

Equation 3

Where \( y_k[\cdot] \) is the \( k^{\text{th}} \) of six onset signals extracted from the filtered audio signals. \( s_{ijk} \) evaluates the similarity between onset signal \( y_k[\cdot] \) and locally generated signal \( g_{ij}[\cdot] \), and it is defined as the inner-product of those two signals. Similarity factor \( S_{ij} \) evaluates the
similarity between all six bands of the audio signal and locally generated signal $g_{ij}[\cdot]$. It is defined as the sum of $s_{ij1}$, $s_{ij2}$, $s_{ij3}$, $s_{ij4}$, $s_{ij5}$ and $s_{ij6}$.

Then, all of the similarity factors are compared and the maximum similarity factor is picked and corresponding index group $(i, j)$ is marked as desired index group $(i_d, j_d)$.

Thus, the period of the music is $T_{id}$ in samples, or $T_{id}/200$ in seconds. The phase of the music is $D_{id}$ in samples, or $2\pi D_{id}/T_{id}$ in radians.

This algorithm requires a large number of calculation, especially when the length of the audio signal is large. In order to reduce the amount of calculation, a simpler method with similar results is presented below. Details about the relation between the original method and the simplified method are discussed in Appendix B.

In this simplified method, a new class of locally generated signals is also generated in two steps. Step one is shown as follows:

$$h_i[n] = \delta[n - T_i] + \delta[n - 2T_i] \quad n = 1, 2, 3, \ldots, 2000$$

$$T_i = 40 + i \quad i = 1, 2, 3 \ldots, 380$$

**Equation 4**

Where $n$ is the index of the signal, and $\delta[\cdot]$ is the discrete impulse. $h_i[\cdot]$ is the $i^{th}$ signal in the group, and $T_i$ is the period of the corresponding signal $h_i[\cdot]$ in samples.

Step two: with respect to each one of signals $h_i[\cdot]$ generated above, a group of signals are generated as follows:

$$h_{ij}[n] = \delta[n - T_i - D_j] + \delta[n - 2T_i - D_j] \quad n = 1, 2, 3 \ldots, 2000$$
Where \( h_{ij}[\cdot] \) is the \( j^{th} \) signal generated from \( h_i[\cdot] \), and \( D_j \) is the corresponding time delay in samples. The number of signals generated from \( h_i[\cdot] \) numerically equals to \( 2000 - 2T_i \). Compared with the class of signals \( g_{ij}[\cdot] \), each of the new signals \( h_{ij}[\cdot] \) consists of two discrete impulses.

Then, the six onset signals extracted from filtered audio signals are used to be compared with the new class of locally generated signals \( h_{ij}[\cdot] \). In this case, the period and phase of the music beats cannot be achieved at the same time. A new similarity factor is defined to find period of the beats first:

\[
D_j = j \quad j = 0,1,2,3 \ldots 2000 - 2T_i \quad \text{Equation 5}
\]

\[
\begin{align*}
\text{Where } h_{ij}[\cdot] \text{ is the } j^{th} \text{ signal generated from } h_i[\cdot], \text{ and } D_j \text{ is the corresponding time delay in samples. The number of signals generated from } h_i[\cdot] \text{ numerically equals to } 2000 - 2T_i. \\
\text{Compared with the class of signals } g_{ij}[\cdot], \text{ each of the new signals } h_{ij}[\cdot] \text{ consists of two discrete impulses.}
\end{align*}
\]

Then, the six onset signals extracted from filtered audio signals are used to be compared with the new class of locally generated signals \( h_{ij}[\cdot] \). In this case, the period and phase of the music beats cannot be achieved at the same time. A new similarity factor is defined to find period of the beats first:

\[
m_{ijk} = \sum_{n=1}^{2000} y_k[n] \cdot h_{ij}[n]
\]

\[
M_{ij} = m_{ij1} + m_{ij2} + m_{ij3} + m_{ij4} + m_{ij5} + m_{ij6}
\]

\[
N_i = \frac{1}{2000-2T_i} \sum_{j=1}^{2000-2T_i} M_{ij}^2
\]

\[
i = 1,2,3 \ldots 380; \quad j = 1,2,3 \ldots 2000 - 2T_i; \quad k = 1,2,3 \ldots 6 \quad \text{Equation 6}
\]

Where \( y_k[\cdot] \) is the \( k^{th} \) of six onset signals extracted from the filtered audio signals. \( m_{ijk} \) evaluates the similarity between onset signal \( y_k[\cdot] \) and locally generated signal \( h_{ij}[\cdot] \). It is defined as the inner-product of those two signals. \( M_{ij} \) evaluates the similarity between all six bands of the audio signal and locally generated signal \( h_{ij}[\cdot] \). Similarity factor \( N_i \)
evaluates the mean of the square of similarity between all six bands of the audio signal and locally generated signals with same period $T_1$.

Then, all of the similarity factors are compared and the maximum similarity factor is picked and corresponding index 'i' is marked as desired index 'i$_d$'. Thus, the period of the music is $T_{i_d}$ in samples, or $T_{i_d}/200$ in seconds.

Since the desired index $i_d$ is known, a group of $M_{i_dj}$ can be calculated by Equation (6) and (8) with index $i = i_d$ and $j = 1,2,3 \ldots 2000 - 2T_1$. The maximum of the group of $M_{i_dj}$ is picked and corresponding index 'j' is marked as desired index $j_d$.

Thus, the phase of the music is the remainder of $D_{i_d}$ divided by $T_{i_d}$, represented as $\text{mod}(D_{i_d}, T_{i_d})$ in samples, or $2\pi \cdot \text{mod}(D_{i_d}, T_{i_d})/T_{i_d}$ in radians.

2.4 Steady State Error

The above procedure is modified in this section in order to reduce the steady state error in the period and phase of the beats. The error is caused because the phase of the beat in the onset signal is not equals to that in the original audio signal.

The difference between the results from the algorithm and the actual value has two components: firstly, when the audio signal is convolved with the half-Hanning window signal, a 0.113 second delay will appear on the envelope signal. The delay occurs due to that the length of the half-Hanning window used in the algorithm. The window has 5000 sample points, and 5000 divided by the sample rate, which is 44100 Hz, is about 0.113 second.
The second part is the differentiator step. Simulation results shows that the phase detected from the onset signal is about 0.03 second earlier than the point where the beat has the largest energy. The differences can be seen in Figure 4 (c) and (e). Since differentiators are used to extracting the onset signal, the phase detected from the algorithm is the start of the beat as mentioned in Section 2.1. However, when human predicts the music beats, one usually predicts the middle of the beats. So this error should be considered and modified.

Considering the two parts of big error, the phase of the beats achieved from last section should be modified by 0.08 sec to overcome the error.

2.5 Music Intensity Analysis

Shiratori et al. [28] have made an assumption that the spectral power of the music increases when the intensity of the music increases. Then, the music intensity can be defined as the power of the music. Similarity factor $N_l$ can be achieved from Equation (9), and the index is chosen to be the desired index $i = i_d$. It can be seen from Equation (9) that $N_l$ has a positive correlation with the power of the music. The difference between $N_l$ and the power of the music is shown in Appendix B. So, the similarity factor $N_l$ is a reasonable approximation to the intensity of a piece of music.
Chapter 3: Simulation

The simulation results of the two-link robot dancing to the music are presented here. It consists of two parts. Part one shows how the algorithm deals with calculation of two signals recorded from two sources: an 80 beats per minute (bpm) metronome and a Latin dance music. Part two gives the simulation result of the robot dance to the beats of the Latin dance music. The model of the robot used in this article is developed in Appendix A.

3.1 Result of the Beat Predicting Algorithm

In this part details about how the algorithm works are given by two examples. One is a piece of music recorded from 80 bpm metronome and another is a piece of 10 seconds duration from a piece of Latin dance music. The audio signals, the envelopes and the extracted onsets are shown in Figure 4.

The onset of signals is extracted from the two pieces of audio signals.
Figure 4 Onset Extraction
Figure 4 shows the onset extraction procedure for an 80 bpm metronome signal with noise (a), (c), (e) and a music example (b) (d) (f). Figure (a) and (b) show the audio waveforms; Figure (c) and (d) show the envelopes; Figure (e) and (f) show the extracted onsets.

The onset signals are convolved to a group of model signals with different periods. The peak of the summed energy is picked out.

Figure 5 (a) shows the output of summed energy from six different bands of an 80 bpm metronome signal. It can be seen from figure that the calculated period located at point 149; under the condition the sample frequency is 200 Hz that it is 80.5 bpm. This result is very close to the real period of the signal which is 80 bpm. The music intensity of the metronome signal is $1.73 \times 10^6$.

Figure 5 (b) shows the summed output of the sum energies from six different bands of the music sample signal. The point 94 has the maximum summed energy, which represent the tempo of the music example is 128 bpm, or 0.47 sec. The music intensity of the Latin music is $3.138 \times 10^7$.

In Figure 5(a), there exists a large pulse, and there is no other pulse. However, in Figure 5(b), there are three smaller pulses other than the largest one and the amplitudes are close. These peaks share a common divisor. These side peaks are caused by eighth notes in the music example. Unfortunately, the algorithm is sensitive to eighth notes and may provide wrong result if strong power occurs at eighth notes.
Figure 5 Tempo estimates using the algorithm.
In Figure 5, the x-axis is labeled samples in the 200 Hz signal and y-axis is the similarity corresponding to different parameter 'a'. (a) Shows the results of an 80 bpm metronome signal with noise and (b) the music sample.

Figure 3-3 (a) shows the summed output of sum energy from six different bands of an 80 bpm metronome signal. The point with largest energy locates at point 843. And modulus 843 by period 149 is 98 in sample, which is 0.49 sec in time domain. The result is modified by minus 0.08 sec, and changed to 0.41 sec. The period and phase of the 80 bpm signal directly extracted from the amplitude figure (Figure 5(a)). This result is close to the time of the first beat in the 80 bpm metronome signal, 0.418 sec and can be seen clearly in Figure 5(a).

Figure 6 (b) shows the summed output of the sum energies from six different bands of the music sample signal. It can be seen that the beat point is 121 in sample. The remainder is 27 samples when 121 samples are divided by the period, which are 94 samples here. So the phase is 0.135 sec. Then the result should be modified by reducing 0.08 sec and changed to the final result that phase of the music sample signal is 0.055 sec.
Figure 6 Phase estimates using the algorithm.

In Figure 6, the x-axis is labeled samples in the 200 Hz signal and y-axis is the similarity corresponding to different parameter 'a'. (a) Shows the results of an 80 bpm metronome signal with noise and (b) the music sample.
So far, the tempo and phase of the beats and intensity of two pieces of music have been detected. The rhythm of the 80 bpm metronome signal is 80.5 bpm and the phase of the beat at 0.41 sec. The rhythm of the Latin music sample is 128 bpm and the phase of the beat at 0.055 sec.

The intensity of the 80 bpm metronome signal with noise and the music sample can be achieved from Figure 5. The music intensity for the 80 bpm metronome signal with noise is $1.73 \times 10^6$, and that for the music sample is $3.138 \times 10^7$.

3.2 Robot Choreography

Two parts of the motion features are used here: keyframe and intensity [28]. Keyframe represents the robot is synchronized motion to the beats of the music. Intensity represents the amplitude of the synchronized movement.

Figure 7 shows the Two-Link robot model moves to the reference input. When a beat comes, the robot starts to move from one position to another position and remains in that place. Then it moves back when the second beats comes. Thus, the robot moves forward and backward periodically to the beats of the music.
Figure 7 the Two-Link robot model moves to the reference input $\theta_1$ and $\theta_2$.

When the robot dances, the rhythm of dance motions should be synchronized the beat of the music. Besides, the range of the movement, or the amplitude of the desired angle can be tuned according to the music intensity. The reference input in the simulation is generated as follows.

The predicted beat positions, calculated previously are used to module the input to the Two-Link robot. First, a sinusoid signal is generated first in the following form related to period and phase of the beats.
\[ y_{\text{sin}}(t) = \sin (2\pi \cdot a \cdot t + b) \quad \text{Equation 7} \]

Where \( a \) is the period of the sinusoid signal, and it is set to be twice the period of music beats; \( b \) is the phase of the sinusoid signal, and it is set to be half the phase of music beats.

Then the reference can be generated related to the sinusoid function that:

\[ y_{\text{ref}}(t) = \begin{cases} \theta_{p1} & \text{if } y_{\text{sin}}(t) \geq 0 \\ \theta_{p2} & \text{if } y_{\text{sin}}(t) < 0 \end{cases} \quad \text{Equation 8} \]

Where \( \theta_{1} = [\sigma], \theta_{2} = [0] \).

In order to have the motion intensity component synchronize to the music intensity, the desired \( \theta \) in the reference input is set as follow:

\[ \sigma = \frac{1}{7.5} \log_{10} N_{i_d} \quad \text{Equation 9} \]

In this form, the reference input changes to \( \theta_{1} \) when a beat comes, and remains until the second beat comes. Then it turns into \( \theta_{1} \) when the second beat comes.
Figure 8 Flowchart of the system

Figure 8 shows flowchart of the system. When the audio signal comes, the beat prediction algorithm is used to calculate and predict the future beats. Then reference generator generates the reference signal of the robot control system according to the predicted beats. The robot control system will have out robot move to the reference input with little steady state error.
Chapter 4: Conclusions and Future Work

A music algorithm was developed that can successfully predict the beat and detect the intensity of a piece of music. The predicted beat position and music intensity from the algorithm can be used to generate a reference input signal. A two-link robot was presented that represents a human arm-forearm. The dynamics and performance of this robot in a dance were formulated and simulated in Matlab. A feed-forward and feedback control system was developed and controlled by reference input signals. The feed-forward and feedback gains were the result of co-activation of pairs of agonist antagonist actuators. The velocity feedbacks were the result of viscosity of the contracting muscles—the Hill effect [8].

The robot model represents a step in creating an intelligent robot dancer that could synthesize dance movements and dance to music. To achieve this, future research is needed in several directions: a robot model based on human dynamics in three dimensions should be developed, the beat prediction method should be further developed to do what humans do and extend to real-time systems.
References


dynamics and stability of multi rigid body systems*. Automatic Control, Vol. 12,
pp. 64–69, 2002.


Control of High-dimensional Humanoids*. Proceedings of the IEEE 2012
International Conference on Robotic and Automation (ICRA), Minneapolis, 2012.


Hasegawa, H. Tsujino, *A robot uses its own microphone to synchronize its steps
to musical beats while scatting and singing*. Intelligent Robots and Systems, pp.
2459-2464, 2008.


Figure 9 Two-Link Model of the Robot

Figure 9 shows the sketch map of the Two-Link robot. It can be seen that the robot has two links in two Dimensions. And the robot has two degree of freedom. Which means the
robot can move to any desired position when angles $\theta_1$ and $\theta_2$ are properly tuned. In order to control the angles, suppose that there are two torques $u_1$ and $u_2$ each applied at connection points of two links.

Using Newton-Euler equations we can get the following equations corresponding to this model.

\[
\begin{align*}
I_1 &= J_1 + m_1 k_1^2 \\
I_2 &= J_2 + m_2 k_2^2
\end{align*}
\]  
Equation 10

\[
\begin{align*}
I_1 \ddot{\theta}_1 &= u_1 - m_1 g k_1 \sin \theta_1 - F_1 l_1 \cos \theta_1 + G_1 l_1 \sin \theta_1 \\
I_2 \ddot{\theta}_2 &= u_2 - m_2 g k_2 \sin \theta_2 - F_2 l_2 \cos \theta_2 + G_2 l_2 \sin \theta_2
\end{align*}
\]  
Equation 11

\[
\begin{align*}
m_1 \ddot{x}_1 &= F_0 - F_1 \\
m_1 \ddot{x}_2 &= F_1 - F_2
\end{align*}
\]  
Equation 12

\[
\begin{align*}
m_1 \ddot{y}_1 &= G_0 - G_1 + m_1 g \\
m_1 \ddot{y}_2 &= G_1 - G_2 + m_2 g
\end{align*}
\]  
Equation 13

\[
\begin{align*}
x_1 &= k_1 \sin \theta_1 \\
y_1 &= k_1 \cos \theta_1
\end{align*}
\]  
Equation 14

\[
\begin{align*}
x_2 &= l_1 \sin \theta_1 + k_2 \sin \theta_2 \\
y_2 &= l_1 \cos \theta_1 + k_2 \cos \theta_2
\end{align*}
\]  
Equation 15
Where $m_1$ and $m_2$ are masses of two links, $l_1$ and $l_2$ are lengths of two links, $k_1$ and $k_2$ are moments of inertias of two links, $k_1$ and $k_2$ are distances between center of mass of the links and joint points.

Then the equations above are summarized and the differential equation corresponding to $\theta_1$ and $\theta_2$ can be achieved.

$$J(\Theta)\ddot{\Theta} + B(\Theta, \dot{\Theta})\dot{\Theta} + gG(\Theta) = W \cdot U \quad \text{Equation 16}$$

Where $\Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$, $W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$.

$$J(\Theta) = \begin{pmatrix} J_1 + m_1k_1^2 + m_2l_1^2 & m_1l_1k_2 \cos(\theta_1 - \theta_2) \\ m_1l_1k_2 \cos(\theta_1 - \theta_2) & J_2 + m_2k_2^2 \end{pmatrix} \quad \text{Equation 17}$$

$$B(\Theta) = \begin{pmatrix} 0 & m_1l_1k_2 \sin(\theta_1 - \theta_2) \dot{\theta}_2 \\ -m_1l_1k_2 \sin(\theta_1 - \theta_2) \dot{\theta}_1 & 0 \end{pmatrix} \quad \text{Equation 18}$$

$$G(\Theta) = \begin{pmatrix} (m_1k_1 + m_2l_1) \sin \theta_1 \\ m_2k_2 \sin \theta_2 \end{pmatrix} \quad \text{Equation 19}$$

The Two-Link robot model has already be presented and can be used to perform movements.

To better control the Two-Link Robot model, the model need to be linearized first.
Suppose there exist an equilibrium point $\theta_e$ that $\theta_e = \theta - \phi$. Then, we can get the linear state space equation around equilibrium point $\theta_e$, that

$$
\dot{x} = \begin{bmatrix}
0 & I \\
-J^{-1}L & 0
\end{bmatrix} x + \begin{bmatrix}
0 \\
J^{-1}
\end{bmatrix} (u + u(\theta_e))
$$

Equation 20

Where $x = \begin{bmatrix}
\varphi_1 \\
\varphi_2 \\
\vdots \\
\varphi_{l_1}
\end{bmatrix}$.

$$
J = \begin{bmatrix}
J_1 + m_l k_1^2 + m_2 l_2^2 & m_2 l_2 k_2 \cos(\theta_{2e} - \theta_{1e}) \\
m_2 l_2 k_2 \cos(\theta_{2e} - \theta_{1e}) & J_2 + m_2 k_2^2
\end{bmatrix}
$$

Equation 21

$$
L = \begin{bmatrix}
-g(m_1 k_1 + m_2 l_1) \cos \theta_{1e} & 0 \\
0 & -g m_2 k_2 \cos \theta_{2e}
\end{bmatrix}
$$

Equation 22

$$
u(\theta_e) = \begin{bmatrix}
g(m_1 k_1 + m_2 l_1) \sin \theta_{1e} \\
g m_2 k_2 \sin \theta_{2e}
\end{bmatrix}
$$

Equation 23

Here, the linearized the state-space model of our Two-Link robot are presented, a method to control the model is show below.

To control the Two-Link robot, we use degree and speed feedback control and the flowchart is shown in Figure 10 below.
Where \( u = -Kx \), and \( K = [K_1, K_2] \).

Then, this equation is substituted into the state space function:

\[
\dot{x} = \begin{bmatrix}
0 & I \\
-J^{-1}(L + K_1) & -J^{-1}K_2
\end{bmatrix} x + \begin{bmatrix}
0 \\
J^{-1}
\end{bmatrix} u(\theta_e) \tag{Equation 24}
\]

Thus, the Two-Link robot can move to a desired position (in the equation is \( \theta_e \)) in some time and stay in that position by choosing appropriate \( K \).

In this paper, feedback control method is used to control the Two-Link robot. Position feedback and velocity feedback are used in the control system to make sure that state space error is very small and the robot can follow the reference input in a relatively short
time. Using the Two-Link robot model we mentioned in section 2, and set parameters in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>m₁</th>
<th>m₂</th>
<th>l₁</th>
<th>l₂</th>
<th>I₁</th>
<th>I₂</th>
<th>k₁</th>
<th>k₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>8</td>
<td>4</td>
<td>0.41</td>
<td>0.4</td>
<td>0.9</td>
<td>0.11</td>
<td>0.18</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 1 Parameters used in the Two-Link robot model

In the simulation, the equilibrium point is set to be \( \theta_e = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \). In order to get better response characteristics, the pole placement is done and poles set at

\[
p = [-100 \ -100 \ -200 \ -200]
\]

Equation 25

Then the feedback gains are achieved:

\[
K = \begin{bmatrix} 31478 & 6560 & 472 & 98 \\ 6560 & 2208 & 98 & 33 \end{bmatrix}
\]

Equation 26

Suppose that the initial position of the angle \( \theta_i = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \), and that the desired position

\[
\theta_d = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}
\]
Then the Two-Link robot is simulated and the response curve of angle $\theta$ to step input is shown in Figure 11. It can be seen that the system is stable, with no overshoot, and little stable state error which is acceptable.
Appendix B: Relations between Two Beat Perdition Algorithms

Details about the relation between two beat prediction algorithms presented in the paper are discussed in this section. The reason why the similarity factor $N_i$ can be a reasonable approximation to the music intensity is also presented below.

In the first algorithm, a group of similarity factor $S_{ij}$ from Equation (3) are calculated and compared to predict music beats. However, it can be seen that nearly 40,000 similarity factors are calculated in the algorithm.

The simplified algorithm presented in the paper simplified the algorithm in two ways. First, the class of locally generated signals is simpler. Each of the signals only consists of two discrete impulses. Thus, the inner-product calculation in Equation (6) is simplified. Second, the beat frequency and phase are predicted separately. Only nearly 400 similarity factors $N_i$ are calculated and simplified at the same time. Compared with the similarity factor in the former method, $N_i$ evaluates the similarity between the onset audio signal and a group of locally generated signals with same delays.

In the simplified method, $M_{ij}$ in Equation (6) is the same to the similarity factor $S_{ij}$ in the original method if the locally generated signal only consists of two intervals. Thus, it can be seen that in the simplified method, the onset of audio signals are compared to the locally generated signal part by part.

In the simplified method, substituting Equation (4) and (5) into Equation (6), following equation can be achieved:
The similarity factor $N_i$ is the mean square of $m_{ijk}$ with the same period $T_i$. It has a positive correlation with the power of the audio signal. So, it can be regarded as a approximation to the music intensity.

\[
m_{ijk} = \sum_{n=1}^{2000} (y_k[n - T_i - D_j] + y_k[n - 2T_i - D_j])
\]

Equation 27